# Creating and analyzing multilingual parliamentary corpora

**Harmonization of the sub-corpora**

- Source materials are converted into a TEI format with the help of the GROBID suite

**Quality assurance**

- Manual checks, iteratively, fine grains where needed

**Reuse**

- Rich reuse history already available
- Fellow scholars working on parliamentary debates from various disciplines

| Step 1 | Step 2 | Step 3 | Step 4 | Step 5 | Step 6 |
|--------|--------|--------|--------|--------|--------|

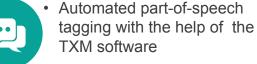**Selection of the data relevant to the research question**

- Public, open data
- English, French, German and parliamentary debates revolving around the European Council meetings

**Creation of the TEI annotation framework**

- TEI annotation framework in two levels: TEI Header (metadata) and on the level of the text (part-of -speech).
- Automated part-of-speech tagging with the help of the TXM software

**Data sharing and publication**

- Corpora and annotation framework: on ORTOLANG
- Papers: on HAL
- Reflections: on a Hypotheses blog